Investigating regulative implications for user-generated content and a design proposal

Garg, Radhika ; Schmitt, Corinna ; Stiller, Burkhard

Abstract: The rapid increase of the Internet connectivity and the data publishing activity, like user-generated content, has lead Internet Service Providers (ISPs) to establish more efficient mechanisms for content delivery, such as caching. Mechanisms such as content-aware-networks and in-network caching reduce network load, server load, and user response time, thus, manage the network. However, caching of content also raises major implications in terms of legal acts and bills (e.g., data privacy, copyright), dealing with access control, validation scheme, and regulations (e.g., contractual obligation, legal restrictions). In general, user-generated content is linked with sensitive information, such as geographical information, medical and financial information, personal identifiable data, photos, videos, and contact information. Therefore, it is essential to secure data and regulate access. The latter, is gained by including access control mechanisms in the data exchange process, where a user requesting data must prove his access rights. Therefore, a user has to show an access ticket, which includes his rights based on legal and regulative implications. In order to secure any kind of data exchange, authentication of each participating communication entity (e.g., content owner, server, and end-user) is essential, which is part of the proposed two-way authentication handshake in this paper that is performed to generate a secure communication channel. The main contribution of this paper is to show that transmission, storage, and usage of user-generated data in caches within the network is manageable within the legal laws on sensitivity, copyright, and privacy. The scope of studying these laws, acts, and policies is restricted to Switzerland (CH), the European Union (EU), and the United States of America (USA). Finally, a solution is presented including access ticketing and two-way authentication mechanisms based on common standards from IP networks.

DOI: https://doi.org/10.1515/pik-2013-0042

Radhika Garg, Corinna Schmitt and Burkhard Stiller*

Investigating Regulative Implications for User-generated Content and a Design Proposal

Abstract: The rapid increase of the Internet connectivity and the data publishing activity, like user-generated content, has lead Internet Service Providers (ISPs) to establish more efficient mechanisms for content delivery, such as caching. Mechanisms such as content-aware-networks and in-network caching reduce network load, server load, and user response time, thus, manage the network. However, caching of content also raises major implications in terms of legal acts and bills (e.g., data privacy, copyright), dealing with access control, validation scheme, and regulations (e.g., contractual obligation, legal restrictions).

In general, user-generated content is linked with sensitive information, such as geographical information, medical and financial information, personal identifiable data, photos, videos, and contact information. Therefore, it is essential to secure data and regulate access. The latter, is gained by including access control mechanisms in the data exchange process, where a user requesting data must prove his access rights. Therefore, a user has to show an access ticket, which includes his rights based on legal and regulative implications. In order to secure any kind of data exchange, authentication of each participating communication entity (e.g., content owner, server, and end-user) is essential, which is part of the proposed two-way authentication handshake in this paper that is performed to generate a secure communication channel.

The main contribution of this paper is to show that transmission, storage, and usage of user-generated data in caches within the network is manageable within the legal laws on sensitivity, copyright, and privacy. The scope of studying these laws, acts, and policies is restricted to Switzerland (CH), the European Union (EU), and the United States of America (USA). Finally, a solution is presented including access ticketing and two-way authentication mechanisms based on common standards from IP networks.

Index Terms: User-generated content, legal policies, content-aware networks, caching, trust, DTLS.

Radhika Garg: E-Mail: garg@ifi.uzh.ch
Corinna Schmitt: E-Mail: schmitt@ifi.uzh.ch
***Burkhard Stiller:** E-Mail: stiller@ifi.uzh.ch

I Introduction

Over the past decades user-generated content has increased manifold. In the USA 82 million people (42.5% of the total of Internet users) created such content in 2008 [40]. A wide variety of applications and Web sites allow publication and viewing of such content, e.g., Flickr [23], Instagram [16], SoundCloud [38], FanFiction.net [15]. These can be categorized into (1) social media Web sites (*e.g.*, Facebook [8] or Twitter [7]) and (2) content sharing Web sites or applications (*e.g.*, YouTube [14] or MySpace [11]). Both categories have in common that the content is available to a broad audience and that circulated content contains sensitive information of the publisher. Personal information through profiles on social networking sites, user behavior, and copyright information through videos, pictures, and text published on the Internet, are important sensitive information linked to user-generated content. Furthermore ISPs try to reduce response time to the request of content by the user, network, and server load by employing caches in the network [27]. Therefore, it is essential that such private and sensitive information is transmitted to users and stored at locations (*e.g.*, caches, servers), which are authenticated and trustworthy. One possibility to bring trustworthiness into the communication way is to perform an authentication process before exchanging data.

Legal regulations, in general, aim to protect private information of the content owner from being propagated in the network without the consent of the owner. User-generated content raises legal questions in terms of intellectual property, defamation, copyright, and privacy rights. Any private information is to be safeguarded from being publicly available, and from malicious attacks. The Swiss and EU Copyright Laws do not allow uploading, transmitting, or copying any copyrighted material without the consent of the content owner. In such a case the legal framework holds the content publisher responsible of the illegal act [31], [10]. The Swiss law permits copying the content, without the consent of content owner, however for private use only. However, copying content in caches located in the network cannot be included in the scope of private use. The World Intellectual Property Organization Copyright Treaty (WCT) is included within the Digital Mil-

lennium Copyright Act (DMCA) in USA [43], [17]. Since the last decade, the European Union also follows this treaty, by the virtue of which, circumventing technical protection measures is also prohibited. Illegal access, transmission, and uploading of any private content are subject to legal consequences of privacy and security breaches [43]. The content transcends many national borders, with the content owner in one country, the content publisher in another, the cache owner in a third, and the user in fourth. When cross-border wrongs (torts) are committed they lead to cross-national litigations [6]. Such litigations can be only completed when the liable party, e.g., content owner, content publisher, cache owner, or ISPs, can be identified appropriately.

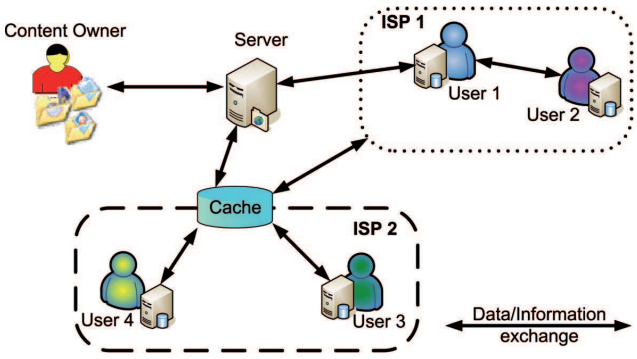

Fig. 1: Assumed Scenario Overview.

The goal of any content distribution network is, therefore, twofold. First, the content has to be delivered with higher efficiency and second, the content delivery process has to be in accordance with legal and regulatory requirements, both determining network and service management functionality. Based on those observations made, this paper addresses the following questions, which to the best of the authors' knowledge have not been addressed so far:

1) What are the legal and regulative implications of using caches in the network?
2) Does a two-way handshake authentication solve the technical requirements to be met?
3) How can a trust-based system on the basis of such a two-way handshake authentication solve legal implications and service management needs of caches in the network?

Hence, this paper presents an approach to authenticate the cached content when (a) it is reused by a user, (b) a different user within the same ISP (owner of the nearest cache to a user) requests the same content for the first time, and (c) a user of a different ISP (not the owner of the nearest cache to the user) requests the same content for the first time. This method can be applied to cache-based content delivery networks, to authenticate users and the content before it is transferred via the cache. This authentication is done to ensure that only legitimate users, as per the legal requirements, have access to the content in caches. This paper concentrates on user-generated content, because (1) such content is heavily cached by content distribution networks, and (2) if monitored and attacked, private and sensitive information about users publishing and consuming (especially viewing and downloading) such content can be retrieved. Due to the main use of caches in cases of delivering user-generated data, the proposal of this work on establishing a trust-based system for authenticating entities exchanging content provides the most beneficial outcomes in terms of fulfilling legal, service management, and network management requirements. Figure 1 illustrates the aforementioned scenario (cf. Section IV).

Today, the security model of the Internet is based on authentication for connection end points [26]. For caching in content-centric networks the content object authentication is required [18]. This means, once an object leaves the original server, its identification has to be still verifiable at any in-network storing locations. This will help to overcome various problems of caching, such as that of copyright and privacy. The trust-based two-way authentication handshake is a method by which the technical benefits of caches, *e.g.*, efficient content delivery, and reduced network load, can be achieved along with fulfilling the regulatory requirement of safeguarding the private information linked with the content. This means access to information is only granted, when it leads to none of these possible law infringements, *e.g.*, copyright infringement, security, and privacy breach. Also, serving copies of content from caches complicates rewarding of benefits in terms of monetary and non-monetary incentives to the content owner or publisher. Such a mechanism of authenticating every access to content in caches can also safeguard such interests.

The remainder of this paper is organized as follows. Section II outlines basic terminology, followed by related work in Section III. Discussion on legal and regulative constraints for caching is presented in Section IV. The scenario assumed is described in Section V and a brief characterization of the proposed solution, including access ticketing and authentication process, is presented. This section also includes the proof of concept of the recommended solution. Finally, Section VII concludes the work and addresses future work.

II Terminology

Content is information that provides value to the end-user in some context. User-generated content means consumer-generated digital content. Organization for Economic Co-operation and Development (OECD) has defined three schools of user-generated content under which, the content should be published in some context, *e.g.*, on a social networking site, a content owner must add own value to the work, and such content is created outside any professional routines and practices [44].

A *stakeholder* can be defined as any kind of entity, who has an interest in the process of generating, transmitting, and using user-generated content.

A *content owner* creates and owns the content. He can also gain benefits in terms of implicit (*e.g.*, social status) and explicit (*e.g.*, monetary) incentives.

A *content publisher* disseminates the content to a wider audience with a monetary benefit. A content publisher is every entity, which wants to publish data (*e.g.*, content owner). *Cache owners* are those stakeholders in a content delivery system, who own the in-network storing locations (*e.g.*,servers or routers) -termed *caches* -that are strategically chosen by an ISP. In comparison to content publishers, entities exist in the network that want to access data, which are called *content subscribers* (*e.g.*, such as (end-)users). Caches, servers, and ISPs are special entities, because they can act either as a content publisher or a subscriber depending on the performed role in the communication way.

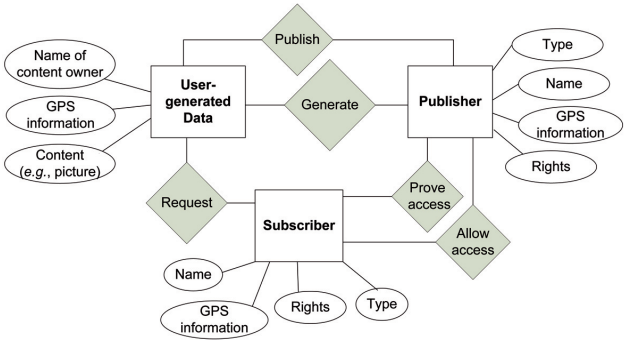

Fig. 2: ER Diagram for Stakeholders.

Figure 2 summarizes all relationships between different stakeholders in an entity-relationship (ER) diagram, where the attribute *type* stands for (end-)user, ISP, server, or routers determining the concrete entity.

III Related work

In Content-centric Networks (CCN) content is transmitted in terms of named objects rather than in a location-dependent manner [18]. In CCNs caching mechanisms are deployed at the network layer. This leads to a major reduction in content retrieval time and bandwidth consumption [20], [18]. Transparent Caching (TC) involves the deployment of intelligent caches in the network to deliver the content fast and more efficiently [29]. This type of caching does not require ISPs to form a contract with the content owner in order to distribute or copy the original content, as it introduces no additional access point, which might lead to security breach, and also the final delivery confirmation is always done by the original server [29]. The status quo of research on establishing cache-based content delivery systems together with taking measures to prevent legal consequences is low. This is because research only concentrate on improving the efficiency of such systems and neglect to incorporate the legal requirements. Researchers have attempted to improve privacy [1], [21] and security [36], [20], [18] issues of CCNs. However, the past research fails to adapt the technology to safeguard the rights of the content publisher in terms of copyright, leakage of sensitive private information, and proper monetary and non-monetary incentives. Thus, Section IV explores the key legal aspects of such a technical approach.

Caches are susceptible to attacks, which may spoil the content or retrieve sensitive information of the publisher or the user [13]. TC, which can be used without the end user and content publisher being aware of, is even more fatal [29]. As a consequence, it is essential to establish a trust-based mechanism for authenticating any request for accessing the content stored in the caches.

Even though researchers acknowledge such implications [39], not much research is done from the perspective of caching for user-generated content, which does not lead to any legal implications. A high level study has been done for analyzing legal implications of caching in peer-to-peer networks [33]. Google caches provide the possibility for the user to bypass the registration process on a Website, hence, making it more susceptible to privacy and security breaches [42]. Privacy attacks on cached content can be based on concept of monitoring 'access to specific content objects by another users connected to same cache' [21]. This means that the content cached in the network can be a source of private information related to content viewer also (*e.g.*, usage patterns or preferences). Adding random delays for new requests and routers will make the identification of a cache location by attackers difficult [24], which

in turn will make the retrieval of private information of content viewer more complicated.

Upon summarizing existing work it can be pointed out that its focus is on efficiency improvements for content delivery networks, however, typically it lacks the consideration of legal requirements for any content access combined with required authentication means.

IV Legal and regulative implications

The understanding of caching both from CCNs and the TC's point of view lead to implications in terms of data protection, privacy, security, copyright. As mentioned before, monitoring usage of cached content of any content publishing platform on the cache can reveal patterns of end-user behavior [24]. Also, traces of cached content can reveal private information about content owners, such as personal profile details. Caching entities located within a network (*e.g.*, routers, in-network servers, or users) can only be used, if they are trustworthy. This is essential, as it ensures that only those entities have access to the data, which do not violate any legal rules. By law caching entities are not allowed to modify the content in any way [10], [17], [43]. However, measures should be taken to ensure that this holds good. For example, Jet Stream, a content delivery system, identifies that transcoding content in real time to lower bit rates is illegal from the net neutrality perspective [25]. This is in fact illegal, as this implies changing the content. Even when the content publisher has a contract with the content owner to publish content on the Internet, it is forbidden to change a single bit of the content. The content stored in caches are copies of the original content, therefore, having no identifiable link to the original copy. This makes identifying copyrights, benefit distribution to the content owner more difficult. Therefore, it is important to authenticate or encrypt content objects, and to request authentication of communication entities by supporting two-way authentication when establishing a secure communication channel. The recommendation of authentication before any access request is granted ensures that the copy of the content cached is always authenticated with respect to the original content source.

The severity of these concerns can vary depending on the way contracts and caching entity owners' treat licensing of rights. Also, when incentives of stakeholders involved in such a scenario deal with pecuniary benefits, implications become more tangled. This is because each copy of the original content might not be linked to the source of the content and, therefore, identifying the owner of the content may be difficult. If the content owner does not get accurate benefits, it might lead to a dispute, between content owner and content publisher. The most important legal implications of caching in the network from the perspective of protecting the private information and benefits for the content owner includes aspects of copyright infringement, liability of content, and incentives of content distribution.

A Copyright Infringement

Copyright issues related to user-generated content arise in several ways. Copyright liability of the content delivery entities, like ISPs or cache owner, who without modification or copying deliver the content, has exemptions by being called 'safe harbors', see DMCA [17] and the EU Electronic Commerce Directive 2000/31/EC [10]. However, a modification of the content for storing it in caches is not allowed. Also, these laws encourage services, which monitor activities on caches and forbid any illegal activities. However, in-network caching is transparent in nature. Hence, questions like where the data is residing and from where it is has been accessed cannot be answered. As a result liability of infringement cannot be identified. This makes it interesting for attackers to attack caches and retrieve copyright information. This grows complex, when such attack types are established for content in transit. The eligibly of a content publisher and cache owner as passive entities (who have no control and knowledge about the content) is still undefined. Copyright infringement by definition questions the reproduction, distribution, and display of copyrighted content. However, consequences of such infringements can also serve as threats in the field of incentives, *e.g.*, monetary benefits due to copyright royalty. Section 103 of the WIPO Copyright and Performance and Phonograms Treaties Implementation Act strictly forbids the circumvention of the copy measurement systems [17]. Hence, copyrighted material by law in Switzerland, EU, and the USA cannot be copied into caches, without prior knowledge and consent of the owner.

B Privacy Protection

A report by OECD stated that content publishers, who provide a platform for content owners to display content (*e.g.*, media files or textual information) get a 'limited irrevocable, perpetual, non-exclusive, transferable, fully paid-up, worldwide license (with the right to sub license) to use, modify, publicly display, reproduce, and distribute

such content through the particular site' [44]. This occurs as soon as a user signs general terms and conditions with any content publisher, mostly without noticing. The consequences for users of giving such permission are that the content publisher now has a right to utilize benefits, without giving a fair share to the content owner. As these rights are transferable, ISPs or cache owners can get the authority to store and modify the original content anywhere in the network, in order to efficiently deal with user requests. Now without the permission of the content owner, content publishers, ISPs, and cache owners can use content. This means, again the fair share of incentive for the content owner cannot be guaranteed.

The Data Protection Directive 95/46/EC of the EU [9] treats unknown monitoring or profiling of any sort as illegal. This law further defines that any information that could be traced to an individual should be treated as private. Therefore, storing of recent or popular requests from the end-user in caches leads to ISPs being informal eavesdroppers recognizing usage patterns, likings, and disliking of the end-user. For example, any activity on an online music portal, if cached, can depict user behavior (which can also be attributed as private information, as it can trace an individual end-user). Private sensitive state can be associated to content with respect to the circumstances it is requested in [11], [20]. Not only this, but access to router caches allows end-users (who have access to the same router cache) to obtain information about their nearby users' content access patterns [1].

Removal of content can never be guaranteed in case of caching. Even if the content owner decides to delete the content, this content will still exist in the network, since deleting or overwriting the content from all caches is not feasible. Debatin et al. investigated this control loss on published content in Facebook. They found out that Facebook works with thousands of content copies in caches, where deleted content (*e.g.*, picture) is still alive after more than three years, although the content owner had deleted the original source [41]. Such caching features forces the content owner to encrypt content to overcome the risk of stealing 'non-existent' data from caches.

Current content delivery systems concentrate on (1) providing newest content to the end-user in an efficient manner and (2) performing content delivery invisible to the content owner and end-user. However, caches store meta information of stored content and, thus, can provide information about private communication traces, which can be exploited by attackers to compromise privacy of the user [20]. Even if content is encrypted, as performed in current content delivery systems nowadays, information can be leaked from meta information, such as content type, source, time stamp, and size.

C Liability of Content

Liability means responsibilty of content in terms of security, privacy, and legality. The onus of liability for content that is in transit and is stored in caches is under constant debate [30]. As soon as the content leaves the original server the liability becomes unclear. ISPs deny the liability, as they prefer to be categorized as passive entities, which merely transfer the content in a more efficient way with the help of caches to end-users [17], [10]. However, leaking of sensitive information to a user, a group of users, or organizations should be taken care of by cache owners or content delivery systems. For uploading illegal content (*e.g.*, illegal copyrighted work or child pornography) onto the Internet, content owner and content publishers are liable [43]. This is, because the content owner owns content and the content publisher facilitates the transmission.

D Incentive Distribution

Serving end-users' requests from caches is a complicated task, when incentives for the content owners have to be evaluated [2]. This is because if the cache serves copies of the original content, the original server is not aware of the request of the content from the end-user. In order to illustrate this task, consider scenario shown in Figure 1, where a media file is stored in a cache and is requested by an end-user in a particular ISP or the first time. Instead of retrieving the content from the original server for each subsequent request, it will be served from this cache. Assuming that the content displaying Web sites connects advertisements with this media file. When requests are served from caches in the network, the latest copy of the content is not retrieved from the server each time. In such a scenario share of benefits (both, monetary and non-monetary) for the content owner is not guaranteed. This happens, because there is no way to assure that the acknowledgment of re-requests of content is also sent to the original server. It can under no circumstances be guaranteed through formal methods that the retrieved content is the most recent content on the original server [5], because the caches refresh their content periodically (*e.g.*, reference [18]). This content update can happen either from the original source or from caches in the Internet. As a consequence, the content publisher looses his control over the user access to his content. If this is the

case, misuses of the caching capabilities can happen, such as duplication, deletion or manipulation of original content. This potential misuse leads to monetary and credibility losses to content owner and to any type of content publisher.

E Discussion

Legal questions rise when data is cached independent of the performed way. Data can either be cached transparently, which means that no contract exist between content owner and cache owners, or in manner of CCN strategies where a chain of contracts between content owner, provider, cache owner, and subscriber exists. In CCN case, the copying of content based on the name, as introduced in Section III, is a high threat. In comparison, copying and disseminating information from the in-network cache without informing the content owner, as performed in transparent caching, raises questions of privacy and copyright. The most crucial implication of deploying caches in the network from the legal and regulative point of view can be summarized at follows:

Protection: The user-generated content is not only copyrighted material, but also contains private and sensitive information of content owners. Storing the content in caches can lead to privacy breaches. Also, the usage pattern (i.e. popularity or frequency) of such media content, when being monitored stored in a cache, can help malicious entities to reveal user-specific private information. The legal problem is that the copyright information cannot be circulated in the network without the consent of the content owner. Also, any private information has to be protected from leaking in the network.

Incentives: As mentioned before, serving the user request from the nearest cache leads to inaccuracy in benefit returns the content owners should get. Measures should be taken by which it is assured that content publishers and subsequently content owners are informed about each content access (even if it is served from cache). Including acknowledgement messages in data exchange protocols, which are send out automatically every time when content is handled in any way can do this. The legal impact is a dispute that can rise from such a scenario, which a content owner can have with a content publisher when he does not get proper incentive.

Liability: Proper counter measures against security, illegal content monitoring, and privacy breach should be part of liability division between various stakeholders. Liability makes the stakeholder responsible for the damage or loss caused by his actions. Liability identification makes it possible for the court to resolve any disputes, since responsibilities can be traced back.

In order to safeguard the content from illegal access and to protect content owners and end-users from loosing their private information of any sort, a DTLS-based solution (cf. Section V) is proposed. If the cache is accessed via performing an authentication process, it overcomes the problem of leaking copyrighted material to unauthenticated content subscribers. This process also makes it possible, even in the case of copied content on caches, to give legitimate incentives to the content publisher.

This solution makes sure that only authenticated and content subscribers have access to the content in the cache. This would prevent copyright, privacy infringement, and achieve in giving appropriate incentive to the content owner. The drawback of the proposed solution is the latency, which occurs, when authent#Imitpunkt#cation is performed and access tickets must be requested and generated. For this performance the devices must have enough resources, especially memory for key storage and energy for upcoming encryptions and calculations.

Access control to cached data is restricted by users attestation specification, where access rights are defined, or by law restrictions applied to ISP (*e.g.*, no access to content from the US). One solution to control data access can be reached by an authentication of communication partners, which must identify themselves and present a corresponding access ticket to the stakeholder of the data cached. Such an authentication can be achieved by integrating a two-way authentication handshake into the communication process before the data exchange takes place [19] recommends adding access-ticketing solution to the process. This means, that the subscriber must present an access ticket to the publisher in order to prove his legitimation for the requested content. In general, those access tickets include an expiration date stamp, which regulates the acceptance of the ticket. If session resumption occurs, due to a connection loss, it is possible to reuse the access ticket if it is still valid. In this case, the two-way authentication handshake can be shortened (cf. Section V-B).

V Application scenario for data access and exchange

The project FLAMINGO currently investigates theoretically access regulations on user-generated content published in

the Internet [12]. An application scenario is given in the project SmartenIT [37], which develops practical solutions for publishing user-generated content and content access in the Internet.

For example, in the global service mobility scenario by SmartenIT where data, which includes sensitive information, is exchanged via a number of hops over longer distances and perhaps is cached in between. PiCsMu deals with a variation of the aforementioned topic, because it is a file storage and file sharing application [22]. Here sensitive data is represented by file content itself and content owner information. The data is fragmented and encrypted before it is stored on different cloud service providers and can only be accessed by authorized users. Before users access requested data they have to prove their access rights, which is mainly investigated in the FLAMINGO project. Due to the fact that the stored data is encrypted the subscribers must have the correct key in order to decrypt accessed data correctly.

A Scenario Description

Figure 1 illustrates the scenarios for this paper and it is assumed that a content owner wants to publish any kind of data on the Internet in order to make it available for different types of subscribers, like users or application programs. Data is transmitted to a server connected to the cloud and content includes sensitive information regarding the content owner. Several users, who are located in different ISP zones (*e.g.*, located in the USA or Europe), want to access the published data. This is possible in several ways as illustrated in an abstract way in Figure 1:

1) Direct content request from the server as processed by user 1 in ISP1.
2) Indirect content request via an intermediate hop as it is the case for user 2.
3) Content request from an ISP via a nearby cache, which is located between the server and the ISP, as processed by user 3 and 4.
4) Users, such as 1 and 2 located in ISP1, can access the published data using the cache of another ISP (*e.g.*, ISP 2). Although, this can happen only, when ISPs have contracts and have agreed to provide each other access to their caches. If no agreement exists the data cannot be accessed using this specific cache. Therefore, another location must be found where access rights exist (*e.g.*, users in ISP 1 can contact the server – YouTube – directly).

B Recommended Access Regulations

Due to the fact that the content owner always has linked private and sensitive information to published content, the authors of this paper recommend that the access to the cached data must be restricted and the subscriber must be authorized. Therefore, it is a suitable option to work with access ticketing solutions as specified in [28] and [3]. Access ticketing solutions allow for a content subscriber to present a temporary access right for the content owned by the content publisher. The content subscriber receives (cf. Figure 3) the access ticket from an Access Control Server (ACS). The access rights are based on legal rules and regulations that are bound to the content as mentioned in Section. III and are included in the access ticket received from the ACS. For example, a content subscriber might only have the right to access data, which is available in its ISP range (*e.g.*, Europe). This ticket is presented to the content publisher, which has to accept or deny the access request. If the ticket was accepted a data exchange could take place, but might be performed over an unsecured communication channel.pt?>

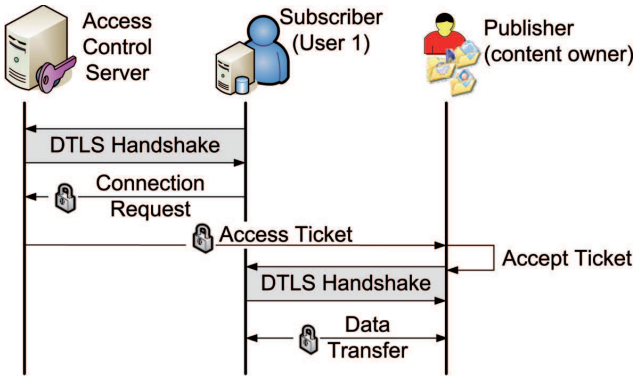

Fig. 3: Process from Generating Access Tickets until Final Data Transfer.

In the next step a DTLS-secured channel is established, where participating communication entities perform a DTLS handshake before exchanging the data itself. In case of the scenario considered in Figure 3 the ACS, the subscriber, and the publisher represent communication entities. Areas marked in grey include the message exchange performed during the DTLS handshake. In general, different authentication options exist for the DTLS handshake depending on the requested security level by entities: Either only one of the communication entities has to authenticate (*e.g.*, server or client) or both entities authenticate each other (termed mutual authentication). The highest level of

security is gained, if both entities perform the mutual authentication by using certificates, which are signed by a trusted third party (*e.g.*, certificate authority) and presented to the other entity.

For a high trustworthiness it is recommended that communication entities perform a two-way authentication handshake as proposed in [19]. The original solution focuses on very constraint hardware, and is transferred here with minor modifications (*e.g.*, message content, messages sizes, or certificate types) to more resource-full devices, like ISP routers. This approach is based on RFC 4347 [32] and assumes that no fragmentation of messages is required, which allows for an UDP support. Figure 3 illustrates that the two-way authentication handshake will be performed twice, if access ticketing is assumed as motivated before. This means for the communication partners different roles occur: The first handshake takes place between the ACS and the content subscriber in order to generate the access ticket. In terms of RFC 4347 the ACS is the server and the content subscriber client. After the access ticket was generated and accepted by the content provider the second handshake takes place, where in terms of RFC 4347 the content provider acts as server.

The DTLS handshake performed includes the following message exchanges based on reference [32]:

1) The client sends a *ClientHello* including a cookie (or the access ticket) to the server.
2) The server has two possibilities to react: If the server is unable to verify the cookie (or the access ticket) received, the server sends back a *HelloVerifyRequest* including the cookie in order to check the aliveness of the client. If the server can handle the handshake latency (*e.g.*, accepting older access tickets by client), the server skips the *HelloVerifyRequest* and sends the *ServerHello* directly to the client.
3) The client sends a *ClientHello* including the supported protocol version and cipher suites.
4) The server answers with a block of messages including a *ServerHello* with the chosen cipher suite and half pre-master secret, a *Certificate* in order to authenticate, a *CertificateRequest* to the client, where client should authenticate itself, and a *ServerHelloDone*.
5) The client answers with a block of messages including a *Certificate*, *ClientKeyEchange*, *CertificateVerify*, *Change-CipherSpec* announcing the cipher suite and keying material negotiated, and closes with a *Finished*.
6) The server concludes the handshake with sensing a *ChangeCipherSpec* and a *Finished* to the client.

The first two message exchanges are optional and can be dropped, if the client and server support session resumption or just perform the TLS handshake. The message Finish is essential to show to the other communication party that the handshake is concluded and the handshake was performed successfully, followed by secure data exchange. In reference [19] it is pointed out that the most trustful handshake can be performed when communication entities include Trusted Platform Modules (TPM). With this add-on the trusted computing functionality can be performed, which means that the trust of a system is based on the hardware and software configuration [4]. In concrete terms this means that a chain of trust is build from the booting of the system until a key generation, resulting in a storage root key. This key is stored in a tamper-proof storage in the TPM and never leaves it. If keys are required for a secure communication, special keys are derived from the storage root key (*e.g.*, signature key, symmetric or asymmetric keys) [4].

For further details of a suitable implementation see [19], [35] on security aspects and [34] on an evaluation.

C Proof of Concept

This solution proposed does not only provide for a security assurance for content transmission, but also has its fundamental basis in mandates provided by legal and regulative organizations across Switzerland, EU, and the USA. In order to illustrate the mapping of the recommended solution to the legal implications explained in Section IV, the scenario assumed and as shown in Figure 1 are as follows: Consider case 4 of the scenario, where user 1 of an ISP zone (here ISP 1) tries to access content, which is stored in a nearby cache owned by another ISP (here ISP 2). As soon as an authentication-based system is integrated it fulfills the legal requirements by (a) checking access rights of the content requesting entity (*e.g.*, end-user or ISP) as soon as copyrighted material from caches is requested, (b) giving a notification to the content owner and/or content publisher whenever the content is being viewed or downloaded by anyone in the network, and (c) preventingillegal access for any sort of content stored in the network, so that malicious users cannot retrieve sensitive information.

The authentication system also sends information to the original server, each time a cached copy of the content is accessed. In such a way incentives of stakeholders are also assured, since a content owner/publisher receives a notification each time when content is accessed.

VI Summary, conclusions, and future work

This paper discusses the impact of legal regulations on user-generated content, which infiuences data access regulations for published data in the Internet. This kind of content is specifically attached with sensitive and private information of both content owner and the end-user. Therefore, caching such content within a network is prone to various legal constraints in terms of privacy, copyright, and data protection. These constraints act as measures to protect the private information linked to the content. This work identifies the protection in terms of service management functionality, especially copyright and privacy issues as a necessity. Caching of content also raises doubts on liability issues and on giving appropriate benefits to the content owner actualizing incentives by an appropriate method. This paper proposes a solution for implementing these legal constraints in practice within the network management. It ensures that (a) the private information is not shared with unauthenticated users, and (b) incentives are accurately given to relevant stakeholders (*e.g.*, content owner, content publisher). The solution proposes to authenticate any entity, which tries to access the content from the cache, by performing a DTLS-based handshake. This access is granted to users who have the right and privilege to view the content.

Concluding, legislative bodies should identify the liability of any inappropriate event due to anomalous access to the data or respective privacy attacks. Implementing laws that identify each stakeholder's liability of privacy protection of the content can do this. In order to take complete advantage of the efficiency of caching in the network, it is necessary to adopt content delivery systems according to the legal framework and requirements. Such systems have to incorporate the privacy protection requirements in the their strategy of storing and transmitting user-generated content via caches. The maximum benefit of content delivery systems can only be achieved when private, sensitive, and personal information of content owner and content subscriber is fully protected.

As part of future work, this proposed solution will be implemented to investigate its effects on the efficiency of a cache-based content delivery based on functional advantages, which were evaluated and discussed in this paper. The overhead due to the authentication is the prime factor that has be monitored and evaluated. Furthermore, a major challenge of this implementation is the translation of legal rules into a machine understandable format. This means that the access will only be granted to those subscribers, who do not violate a legal mandate in terms of privacy and copyright. Translating and implementing these rules for a broad geographical region (e.g., Switzerland, EU, and the USA) is more complicated, as the legal basis and implications becomes more intricate and varied.

As mentioned in Section IV subscribers must prove their access rights and know the key for encryption purposes. But if the subscriber looses its access right, leaves the network or is attacked, it is important to secure the data in the network. Periodically updating the used cipher suites and keying material in the network can do this. Therefore, the projects FLAMINGO and SmartenIT can investigate key management and revocation issues in the future.

VII Acknowledgement

This work was supported partially by the FLAMINGO and the SmartenIT projects, funded by the EU FP7 Program under Contract No. FP7-2012-ICT-318488 and No. FP7-2012-ICT317846, respectively. Finally, thanks are addressed to M. Charalambides for his valuable input concerning legal implications.

References

1 G. Acs, M. Conti, P. Gasti, C. Ghali, and G. Tsudik. Cache Privacy in Named-Data Networking. In *Proceedings of the 33rd International Conference on Distributed Computing Systems*, ICDCS. IEEE, 2013.

2 P. K. Agyapong and M. Sirbu. Economic Incentives in Information-Centric Networking: Implications For Protocol Design and Public Policy. *IEEE Communications Magazine*, 50(12): 18–26, 2012.

3 T. Burroughs, C. Kibbe, R. Smith, G. Bhatia, K. Biswas, P. Encarnacion, N. Kumar, and A. Swaminathan. Oracle Single Sign-On Application Developer's Guide, Part Number A86782-03. Technical Report 3.0.6, ORACLE, 1999.

4 D. Challener, K. Yoder, R. Catherman, D. Safford, and L. V. Doorn. *A Practical Guide to Trusted Computing*. IBM Press, 2008.

5 J. Choi, J. Han, E. Cho, T. Kwon, and Y. Choi. A Survey on Content-oriented Networking for Efficient Content Delivery. *IEEE Communications Magazine*, 49(3): 121–127, 2011.

6 E. Clemons and Y. Chen. Making the Decision to Contract for Cloud Services: Managing the Risk of an Extreme Form of IT Outsourcing. In *Proceedings of the 44th Hawaii International Conference on System Sciences*, HICSS, pages 1–10. IEEE, 2011.

7 S. Dann. Twitter Content Classification. *First Monday -Peer-Reviewed Journal on the Internet*, 12, December 2010.

8 B. Debatin, J. P. Lovejoy, A.-K. Horn, and B. N. Hughes. Facebook and Online Privacy: Attitudes, Behaviors, and Unintended Consequences. *Computer-Mediated Communication*, 15(1): 83–108, 2009.

9 Directive E-commerce. Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of indovidual with regard to the processing of personal data and on the free movement of such data. *Offical Journal*, L 281: 31–50, 1995.

10 Directive E-commerce. Directive 2000/31/EC of the European Parliament and of the Council of 8 June 2000 on Certain Legal Aspects of Information Society Services, in Particular Electronic Commerce, in the Internal Market (Directive on Electronic Commerce). *Offical Journal*, L 178: 1–16, 2000.

11 C. Dwyer, S. R. Hiltz, and K. Passerini. Trust and Privacy Concern Within Social Networking Sites: A Comparison of Facebook and MySpace. In *Proceedings of the Americas Conference on Information Systems*, AMCIS, pages 339–350, 2007.

12 FLAMINGO Consortium. FLAMINGO: Management of the Future Internet. http://www.fp7-flamingo.eu/, August 2013.

13 P. Gasti, G. Tsudik, E. Uzun, and L. Zhang. DoS DDoS in Named-Data Networking. *arXiv preprint arXiv: 1208.0952*, 2012.

14 P. Gill, M. Arlitt, Z. Li, and A. Mahanti. Youtube Traffic Characterization: A View From the Edge. In *Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement*, IMC, pages 15–28. ACM, 2007.

15 K. Hellekson and K. Busse. *Fan Fiction and Fan Communities in the Age of the Internet: New Essays*. McFarland, 2006.

16 Instagram. http://www.instagram.com, August 2013.

17 M. Jackson. The Digital Millennium Copyright Act of 1998: A Proposed Amendment to Accommodate Free Speech. *Communication Law and Policy*, 5(1): 61–92, 2000.

18 V. Jacobson, D. K. Smetters, J. D. Thornton, M. F. Plass, N. H. Briggs, and R. L. Braynard. Networking Named Content. In *Proceedings of the 5th International Conference on Emerging Networking Experiments and Technologies*, CoNEXT, 2009.

19 T. Kothmayr, C. Schmitt, W. Hu, M. Brünig, and G. Carle. DTLS Based Security and Two-Way Authentication for the Internet of Things. *Ad Hoc Networks, Elsevier*, 2013.

20 T. Lauinger, N. Laoutaris, P. Rodriguey, T. Strufe, E. Biersack, and E. Kirda. Technical Report TR-iSecLab-081-001. 42(8), 2012.

21 T. Lauinger, N. Laoutaris, P. Rodriguez, T. Strufe, E. Biersack, and E. Kirda. Privacy Risks in Named Data Networking: What Is The Cost of Performance? *ACM SIGCOMM Computer Communication Review*, 42(5): 54–57, 2012.

22 G. Machado, T. Bocek, M. Ammann, and B. Stiller. A Cloud Storage Overlay to Aggregate Heterogeneous Cloud Services. In *38th IEEE Conference on Local Computer Networks (LCN 2013)*, Sydney, New South Wales, Australia, October 2013.

23 A. Mislove, H. S. Koppula, K. P. Gummadi, P. Druschel, and B. Bhattacharjee. Growth of the Flickr Social Network. In *Proceedings of the 1st Workshop on Online Social Networks*, WOSN, pages 25–30. ACM, 2008.

24 A. Mohaisen, X. Zhang, M. Schuchard, H. Xie, and Y. Kim. Protecting Access Privacy of Cached Contents in Information Centric Networks. In *Proceedings of the 8th ACM SIGSAC Symposium on Information, Computer and Communications Security*, pages 173–178. ACM, 2013.

25 Net Neutrality Technology Value Chain Cooperation. JET STREAM -Cutting Edge Digital Media Logistics, Transparent Internet Caching. http://www.jet-stream.com/blog/transparent-internet-caching/, June 2011.

26 R. Oppliger. Internet Security: Firewalls and Beyond. *ACM Commun#lmitpunkt#cations*, 40(5): 92–102, May 1997.

27 G. Pallis and A. Vakali. Insight and Perspectives for Content Delivery Networks. *ACM Commununications*, 49(1): 101–106, 2006.

28 B. Patel and J. Crowcroft. Ticket Based Service Access for the Mobile User. In *Proceedings of the 3rd Annual ACM/IEEE International Conference on Mobile Computing and Networking*, MobiCom, pages 223–233. ACM, 1997.

29 PeerApp – Transparent Caching Commercial Deployer. http://www.peerapp.com/products/transparentcaching.aspx, August2013.

30 F. Poggi and A.-I. Riviere-Osipov. Legal Analysis of a Single Market for the Information Society. Digital Agenda For Europe – A Europe 2020 Initiative SMART 2007/0037, European Commission, May 2011.

31 Rentsch Partner AG. Copyright Law. http://copyright.ch/, August 2013.

32 E. Rescorla and N. Modadugu. Datagram Transport Layer Security. RFC 4347 (Proposed Standard), Apr. 2006. Obsoleted by RFC 6347, updated by RFC 5746.

33 P. Rodriguez, S.-M. Tan, and C. Gkantsidis. On the Feasibility of Commercial, Legal P2P Content Distribution. *ACM SIGCOMM Computer Communication Review*, 36(1): 75–78, 2006.

34 C. Schmitt. *Secure Data Transmission in Wireless Sensor Networks*. PhD thesis, Technische Universität München, Germany, Department of Computer Science, Network Architectures and Services, July 2013.

35 C. Schmitt, B. Stiller, T. Kothmayr, and W. Hu. DTLS-based Security with two-way Authentication for IoT, IETF Draft. Technical report, October 2013.

36 SecTheory: Networking and Security Infrastructure Provider. http://www.sectheory.com/rfc1918/security/issues.htm, July 2013.

37 B. Stiller, D. Hausheer, T. Hoßfeld: Towards a Socially-Aware Management of New Overlay Application Traffic Combined with Energy Efficiency in the Internet (SmartenIT); in: Alex Galis, Anastasius Gavras (Edts.) "The Future Internet", Lecture Notes in Computer Science, Springer, Berlin Heidelberg, Germany, LNCS Vol. 7858, May 2013, pp 3–15.

38 SoundCloud Limited. SoundCloud. https://soundcloud.com/, August 2013.

39 D. Trossen and A. G. Kostopoulos. Techno-Economic Aspects of Information-Centric Networking. *Journal of Information Policy*, 2, 2012.

40 P. Verna. A Spotlight on UGC Participants. *eMarketer Inc.*, February 2009.

41 R. Waugh. Deleted Facebook photos still viewable THREE YEARS later. *Mail Online -Science & Tech*, February 2012.

42 Wiki for Bypassing Web Registration. http://www.wikihow.com/Bypass-Registration-on-Websites-using-the-Google-Cache, July 2013.

43 WIPO Copyright Treaty. Technical report, World Intellectual Property Organization (WIPO), Collection of Laws for Electronic Access, 1996.

44 S. Wunsch-Vincent and G. Vickery. Participative Web: User Generated Content, Directorate for Science, Technology and Industry. Technical Report JT03225396, OECD Directorate for Science, Technology and Industry, 2007.

Radhika Garg: Department of Informatics, Communication Systems Group, University of Zurich, Binzmühlestrasse 14, CH-8050 Zürich, Switzerland

Burkhard Stiller: Department of Informatics, Communication Systems Group, University of Zurich, Binzmühlestrasse 14, CH-8050 Zürich, Switzerland

Corinna Schmitt: Department of Informatics, Communication Systems Group, University of Zurich, Binzmühlestrasse 14, CH-8050 Zürich, Switzerland